Health Policy 112 (2013) 70–79

Health Policy

journal homepage: www.elsevier.com/locate/healthpol

International comparisons of the technical efficiency of the hospital sector: Panel data analysis of OECD countries using parametric and non-parametric approaches☆☆

Yauheniya Varabyova*, Jonas Schreyögg[1]

Hamburg Center for Health Economics, University of Hamburg, Esplanade 36, D-20354 Hamburg, Germany

ARTICLE INFO

Article history:
Received 8 October 2012
Received in revised form 5 March 2013
Accepted 9 March 2013

Keywords:
International comparison
Hospital sector
Technical efficiency
DEA
SFA

ABSTRACT

There is a growing interest in the cross-country comparisons of the performance of national health care systems. The present work provides a comparison of the technical efficiency of the hospital sector using unbalanced panel data from OECD countries over the period 2000–2009. The estimation of the technical efficiency of the hospital sector is performed using nonparametric data envelopment analysis (DEA) and parametric stochastic frontier analysis (SFA). Internal and external validity of findings is assessed by estimating the Spearman rank correlations between the results obtained in different model specifications. The panel-data analyses using two-step DEA and one-stage SFA show that countries, which have higher health care expenditure per capita, tend to have a more technically efficient hospital sector. Whether the expenditure is financed through private or public sources is not related to the technical efficiency of the hospital sector. On the other hand, the hospital sector in countries with higher income inequality and longer average hospital length of stay is less technically efficient.

1. Introduction

International comparisons of health care system performance have always drawn the attention of different actors in health policy. In many OECD countries, the health care system constitutes the largest service industry, with an average health spending reaching 9.5% of GDP in 2010. In particular, the hospital sector consumes a considerable share (16–48%) of total health expenditure in OECD countries [1]. The analysis of the hospital industry provides an important contribution to the comparison of health care system performance.

Hospital efficiency is one of the key indicators of hospital performance. Although manifold efforts have been undertaken to evaluate and compare hospital efficiency within a particular country, only a few studies have so far conducted a between-country analysis of hospital efficiency utilizing data from two to four countries [2–8]. The scarcity of studies, which investigate between-country hospital efficiency, has been underlined by the lack of suitable data needed for a comprehensive international comparison [5]. The available studies performed their efficiency analyses using nonparametric techniques, such as data envelopment analysis (DEA) and directional distance function, with the only exception of Kittelsen et al. [3],

☆☆ Open Access for this article is made possible by a collaboration between Health Policy and The European Observatory on Health Systems and Policies.

* Corresponding author. Tel.: +49 40 42838 9525; fax: +49 40 42838 8043.
E-mail addresses: yauheniya.varabyova@wiso.uni-hamburg.de (Y. Varabyova), jonas.schreyoegg@wiso.uni-hamburg.de (J. Schreyögg).
[1] Tel.: +49 40 42838 8041; fax: +49 40 42838 8043.

0168-8510
http://dx.doi.org/10.1016/j.healthpol.2013.03.003

who tested the robustness of their results by comparing the effects of environmental factors on efficiency estimated by two-step DEA with parametric stochastic frontier analysis (SFA).

The analysis of the efficiency of health production at country level in general, without the focus on the hospital sector, using parametric and nonparametric approaches has been attempted previously [9–18]. Most of these studies analyze how efficient countries are in producing health care outcomes, e.g., life expectancy. Färe et al. [9] presented a nonparametric model most closely related to efficiency studies in the hospital sector, which focused on the production of intermediate hospital outputs over the period from 1974 to 1989. However, their inputs did not completely discriminate between the inpatient and outpatient sectors and their output variables did not include any control for case-mix differences.

These shortcomings provided the motivation for the present comparative study of the hospital sector efficiency in OECD countries, which estimates technical efficiency with parametric and nonparametric techniques across a number of models using aggregate hospital data at country level and controlling for the differences in case severity. Furthermore, the estimates of hospital efficiency are regressed on a set of environmental variables to analyze the factors associated with higher levels of efficiency using unbalanced panel data from 2000 to 2009. The purpose of this study is to expand the previous attempts to compare the efficiency of the hospital sector and to examine the environmental factors associated with the cross-country differences in hospital sector efficiency. Moreover, the study aims to enhance the understanding of the use of efficiency methods for the purpose of international comparison of the hospital sector performance.

2. Methods

In the definition of efficiency, the distinction should be made between technical and price (allocative) efficiency measures, which together comprise the overall (economic) efficiency, and scale efficiency [19,20]. This study analyzes only one facet of efficiency, namely, technical efficiency. This decision is necessitated both by theoretical and data considerations, as international hospital cost statistics, needed for the estimation of price (allocative) efficiency, are not available and scale properties do not lend themselves to estimation at country level [3].

Frontier analysis has firmly established itself as a reliable method for the assessment of efficiency. However, a wide variety of frontier models are available with few theoretical or statistical criteria to judge between them. This study compares nonparametric data envelopment analysis (DEA) with parametric stochastic frontier analysis (SFA).

2.1. Data envelopment analysis (DEA)

Around 80% of frontier efficiency analyses published up to mid-2006 employ nonparametric techniques, the principal application being DEA [21]. The popularity of DEA is underlined by its convenient qualities. DEA allows considering multiple inputs and outputs simultaneously and requires no a priori assumptions about the functional form of the production frontier. For detailed technical descriptions of DEA the reader is referred to a number of sources [20,22–25], and only a brief discussion of relevant aspects is given here.

One important assumption to make when performing DEA is whether to use an input- or output-orientation. An input-oriented model holds the current level of output constant and minimizes inputs, whereas an output-oriented model maximizes output keeping the amount of inputs constant. Farrell [19] did not specify a formal definition of the contemporary “Farrell measure” of the technical efficiency of production and did not standardize the two different measures of technical efficiency [26,27]. Deprins and Simar [26] defined input technical efficiency as a measure between zero and one, whereas output technical efficiency as a measure greater than one.

Another important theoretical assumption in DEA is whether to apply constant or variable returns to scale. The first nonparametric models for efficiency estimation by Charnes et al. [28] assumed constant returns to scale (CRS). Later on, Banker et al. [29] incorporated variable returns to scale (VRS) to account for firms, which do not operate at their optimal scale. If ratio data is used in DEA, which is the case in the present study, then for technical reasons the model with the VRS constraint is applied, even though ratio data imply CRS [30].

Nonparametric efficiency measures have been criticized for lacking a statistical basis. DEA is a data-driven approach, where observations are produced by an implicit data-generating process (DGP). Exploring the underlying DGP provides the means to analyze the sensitivity of estimated efficiency scores to sampling variation. Owing to the works of Simar and Wilson [31–33], bootstrapping the DEA efficiency scores has gained in popularity. The idea of bootstrap lies in repeated simulations of the DGP, creating a new dataset of the original size. The original estimator is applied to each simulated sample, producing estimates that imitate the sampling distribution of the original estimator. The bootstrap procedure allows deriving the statistical properties of efficiency scores, by estimating the bias and variance and by constructing confidence intervals.

All efficiency scores obtained in the first stage of DEA and presented in this study are corrected for the bootstrap bias. The DEA input-oriented efficiency estimate is biased upward, because the minimization problem involves a smaller set of decision-making units than the true set. With the help of bootstrap, the original estimates can be corrected downward for the estimated bias value. Alternatively, the application of bootstrap to the output-oriented efficiency case implies correcting the estimated efficiency upward for the bias. This study uses a conventional amount of 2000 bootstrap replications. In the second stage of DEA, the bias-corrected efficiency scores obtained from the cross-sectional estimations for years 2000–2009 are regressed on a set of explanatory variables using the truncated regression model. Truncated regression allows taking advantage of the bootstrap procedure and performs well in terms of confidence intervals coverage [33].

2.2. Stochastic frontier analysis (SFA)

SFA was independently developed by Aigner et al. [34] and Meeusen and van den Broeck [35]. SFA decomposes the error term into two components. One part represents random events outside the decision-making unit's control and the other part is a non-negative term capturing inefficiency. SFA is a parametric technique, which requires assumptions about the functional form of the production function and the distribution of the error terms. In this study, the single-output Cobb–Douglass frontier production function is estimated. Initially, the flexible translog model was also considered. However, the translog does not fit our data, as the model does not converge after 110 iterations and the input terms in the production function are insignificant and some are negative. In contrast, the Cobb–Douglas model provides an excellent fit. It has been shown in simulation studies that a mis-specified translog function performs rather poorly despite its flexibility if the sample size is small [36].

In the cross-sectional SFA models, used in this study to judge the validity of the DEA efficiency estimates, inefficiency terms are assumed to be half-normal distributed. Other hospital studies have employed half-normal, exponential, and truncated distributions. The lack of a priori justification for the use of any particular distribution has been a general criticism of SFA. However, the research has shown that different distributions have a very small impact on the estimates of inefficiency [10,37]. In our study, the correlation between the estimates of inefficiency using exponential and half-normal distributions is very strong (0.99), whereas the estimation based on truncated distribution does not achieve convergence.

The panel data models of SFA distinguish two approaches, concerning the assumption of whether or not efficiency changes over time. Time-invariant efficiency models assume efficiency to be constant over time [38–41]. This assumption is not particularly plausible in the setting where data are collected over long periods or when external periodic influences are expected to affect efficiency [20]. Other approaches assume time-varying efficiency, whereby the general pattern of efficiency change should be modeled. Several approaches to modeling time-dependency in efficiency have been suggested [42–44]. This study uses the approach of Battese and Coelli [45]. In this model, the inefficiency term is defined by the truncation at zero and is directly affected by country-specific factors.

2.3. Validity testing

While the real advantage of DEA is its ability to estimate a multiple input, multiple output model, it is often advisable in a single-output model to use parametric techniques [25]. When the researcher does opt for the nonparametric estimation, two important considerations should be taken into account. First, DEA imposes a strong assumption of no error in the data, which is often quite unrealistic. Second, the nonparametric nature of the analysis creates difficulties in applying the standard diagnostic tests to evaluate the results of DEA [46]. The uncertainty surrounding nonparametric estimates is mitigated by applying the bootstrap procedure, which provides the means to infer the statistical properties of the estimated efficiency scores and to produce the bias-corrected estimates [31,32]. Moreover, the bootstrap procedure allows testing restrictions in nonparametric efficiency models, such as the relevance of inputs or outputs, aggregation of inputs or outputs [47], and returns to scale [48].

The validity of findings can be classified into internal validity, which concerns the stability of the results to changes in the used methods, and external validity, which addresses the applicability of the results in a more general setting [25]. Internal validity can be tested by examining the changes in the efficiency estimates when different combinations of input and output variables are used [46,49] or when different methods are applied within the same dataset, such as input- versus output-orientation [12]. External validity can be tested as the consistency over time [25,50] or by comparing the efficiency scores estimated by DEA with the efficiencies derived by SFA using the same set of input and output variables [46,50–53].

The statistical data analyses are performed using R. Specifically, package 'Benchmarking' [54] is applied for DEA and package 'frontier' [55] is used for SFA.

3. Variables and data specification

Data for the analysis are obtained from the Organisation for Economic Co-operation and Development (OECD) Health Data 2012, which is the broadest source of comparable statistics on diverse health systems across OECD countries. The sources and methods of data collection are described in detail in the OECD documentation [56]. In this study, data for the period 2000–2009 are used. Due to the fact that countries are not uniform in their reporting practices and not all variables are recorded each year, a slight adjustment of OECD data is indispensable and is common in OECD studies (e.g., [12,13]). In this study, linear interpolation is applied to impute missing values in the time-series for particular countries, which basically means that some of the gaps are filled with average estimates (but no more than for a two-year period). Chili, Mexico, and Sweden are completely excluded from all estimations due to missing variables, resulting in a final sample of 31 OECD countries. The definitions of variables used in this study and summary statistics for 2007, which is the year with the most available data points, are presented in Table 1.

3.1. Output variables

3.1.1. Discharges

The ideal measure of final output in hospital care would be some measure of the health gain of individual patients [21]. However, these data are not readily available, which prompts researchers to use some form of an intermediate output in their analyses, which usually involves inpatient days [4,6] or discharges [57]. Discharges are argued to be a better output measure than inpatient days, as unnecessary inpatient days at the margin for a hospital episode might falsely indicate high efficiency [58].

Table 1
Descriptive statistics (year 2007).

Variable	Definition	Measurement	Mean	SD
Input variables				
Beds	Total hospital beds	Density per 1000 population	5.49	2.28
Employment	Total hospital employment	Density per 1000 population (head counts)	15.40	5.22
Physicians	Physicians employed in hospitals	Density per 1000 population (head counts)	1.76	0.58
Nurses	Professional nurses and midwives employed in hospitals	Density per 1000 population (head counts)	4.51	2.13
Output variables				
Discharges	Discharge rates by diagnostic categories aggregated by case severity	Density per 1000 population	174.88	53.55
Mortality	1 – average in-hospital mortality rate	Rate per 100 patients	0.90	0.03
Environmental variables				
Expenditure	Health expenditure	Per capita thousands US$ PPP	3.02	1.37
		% of GDP	8.37	1.94
Hospital expenditure	Hospital expenditure	Per capita thousands US$ PPP	1.05	0.47
		% of GDP	3.08	0.72
Private expenditure	Private sector health expenditure	% of current expenditure on health from private sector	27.24	9.19
Inequality	Income inequality	Gini coefficient after taxes and transfers for the mid-2000s	0.31	0.05
Hospital density	Total hospitals	Density per million population	32.19	17.00
Public hospitals	Publicly owned hospitals	% of total hospitals	0.54	0.25
Education	Population with upper secondary education	Density per 1000 population	12.18	2.41
Length of stay	Average length of stay: in-patient care	Days	9.21	5.40
Population over 65	Population: 65 years old and over	% of total population	14.84	3.30
Life expectancy	Life expectancy at birth	Years	79.03	2.62
Infant mortality	Infant mortality	Deaths per 1000 live births	4.34	2.45
Full-time employment	Incidence of full-time employment	The share of full-time employment in the economy	85.25	7.45

Source: OECD Health Data 2012, OECD Labor 2012.

OECD relies on International Shortlist for Hospital Morbidity Tabulation (ISHMT) in developing a list of diagnostic categories consisting of groups defined by both ICD-9 and ICD-10 codes, which allows comparisons of discharges and length of stay (LOS) by diagnostic category between countries using different International Classification of Diseases (ICD) revisions [59]. As this list has been considered to be too extensive for the purpose of our analysis, the shortlist containing 20 sections of diagnostic categories is used.

The discharges by diagnostic category are aggregated into a weighted measure of total discharges on the basis of case severity using the information on LOS by diagnostic category. A strong assumption to such aggregation is that LOS represents a resource usage implied by a particular illness severity. The aggregation procedure follows the methodology of Herr [60], who developed and applied it to hospital-level data.

First, the global LOS measure is calculated by adding the duration of all stays in days across all discharge categories in all countries and dividing it by the absolute number of all discharges. Second, from this information the index of discharge weights is constructed as a ratio of LOS in a particular disease category and dividing it by the global LOS measure. An index smaller (bigger) than one represents a discharge category, which involves a shorter (longer) than average LOS. The final measure of aggregated discharges adjusted by case severity is then created by multiplying the discharges by diagnostic categories by the index of discharge weights and summarizing the adjusted discharges for a given country. Note that the measure of LOS as applied here is not used to rate the countries but rather to rate the discharge categories, and does not, thereby, lead to the problem of overestimating outputs in countries, which encourage longer LOS.

3.1.2. Mortality

Discharges represent an activity-based measure of output and as such, all else being equal, rate the countries that treat more patients as more efficient [20]. An activity-based analysis considers two countries with the same number of discharges as equivalent, even though patients are more likely to die if treated in one country than another.

In-hospital average mortality rate has been used in recent studies as an additional output variable to control for the potential tradeoff between inefficiency and mortality [61,62]. OECD provides statistics on in-hospital mortality following acute myocardial infarction (AMI) and stroke. Although this measure has limitations, as it only captures a small proportion of all in-hospital deaths, AMI and stroke mortality rates do represent a good proxy measure for hospital quality as they encompass effective medical interventions and timely and coordinated treatment of patients. Consequently, AMI and stroke mortality rates have been employed for hospital benchmarking within and between OECD countries. Beyond the quality of hospital care, AMI and stroke mortality can be influenced by differences in the recording of hospital transfers, average length of stay, and the severity of AMI or stroke, among other factors [56].

In this study, the admission-based 30 day age–sex standardized average in-hospital mortality rate is constructed across three conditions, namely, AMI, hemorrhagic stroke, and ischemic stroke, while treating each condition with equal weight.

3.2. Input variables

3.2.1. Beds

The number of hospital beds represents a measure of the resources, which are available for providing services to inpatients in hospitals. Moreover, Mobley and Magnussen [6] argue that the variable beds can be treated as representing the variation in service technology among different hospitals. As such, the number of beds is conventionally used as an approximation for the capital and technology input in a within-country hospital comparison [25,58,60] as well as in an international context [2,6].

3.2.2. Hospital employment

Labor input in hospital efficiency studies is often disaggregated by skill level [2,6,7]. Although the OECD dataset disaggregates the measure of total hospital employment into six different categories, only physicians employed in hospitals and nurses employed in hospitals provide enough data for analysis. The measurement of hospital employment involves head counts rather than full time equivalents, as most of the data are available in head count units. For a few countries, which provide data only for full time equivalent employment, the conversion into head count units is executed on the basis of the average ratio of head counts to full time equivalents derived from countries, which provide both measures.

3.3. Environmental variables

3.3.1. Health care expenditure

Policies and institutions are designed to influence the performance of health care while containing costs. Evans et al. [63] found health system efficiency to be related to health care expenditure per capita for a sample of 191 countries. This study investigates whether health care expenditure per capita is associated with the hospital sector efficiency in OECD countries.

3.3.2. Financing of health care

Although all OECD countries use a mixed way of financing their health care, they use public and private sources to a various extent. In order to control for the effect of the source of financing, the percentage of private financing in total health care expenditure is used. A high level of private spending may lead to regressive health care financing and thus cause inequalities in the access to health care services [6,64].

3.3.3. Income inequality

The differences in income distribution across countries are measured by the Gini coefficient for the mid-2000s. The coefficient takes the value from 0 to 1, with higher values representing a greater degree of inequality. Previous research has found that inequality might affect health status and health care efficiency at the international [12,65,66] as well as at the regional level [67].

3.3.4. Market influences

In order to account for market influences, the measure of hospital density as total hospitals in a country per million population is included [68]. Higher density of hospitals in a country can potentially lead to competitive pressures as well as a faster spread of technology and efficient managerial practices. Alternatively, the number of hospitals per million population can also be capturing economies of scale, as our measure of hospital density can be interpreted as a proxy for the inverse of the number of patients (in millions) per hospital.

3.3.5. Education

Education has proved to be a key contributing factor to population health status in empirical studies, as education influences many of the decisions, which determine the quality of life and mortality rates [12,69]. In this study, education is defined as the population density with upper secondary education.

3.3.6. Length of stay

OECD countries differ considerably from each other in their average hospital lengths of stay (from 4 to 5 days in Israel, Denmark, and Turkey to 34 days in Japan). Certain remuneration systems, e.g., per diems, might create incentives for hospital managers to extend the length of stay in order to meet their budgets [7,70]. Differences in length of stay may explain differences in efficiency.

3.3.7. Health status

Life expectancy and infant mortality rates are included to control for the heterogeneity in population health status, as suggested in previous research [6]. Moreover, the infant mortality rate might serve as an indicator of the quality of prenatal care, whereas the life expectancy rate characterizes the quality of elderly care [6].

3.3.8. Patient mix

Most hospital studies include some measure of patient mix, in particular, the proportion of elderly patients [58,60,71,72]. The percentage of population aged over 65 is included as an additional control for case-mix differences.

3.3.9. Full-time employment

Since hospital employment in this study is measured in head counts and not in full time equivalent (FTE) units,

Table 2
Model specifications.

	DEA[a]									SFA		
	m1	m2	m3	m4	m5	m6	m7	m8	m9	s1	s2	s3
Input[b] variables												
Beds	X	X	X	X	X	X	X	X	X	X	X	X
Employment	X			X			X			X		
Physicians		X	X		X	X		X	X		X	X
Nurses			X			X			X			X
Output variables												
Discharges	X	X	X	X	X	X	X	X	X	X	X	X
Mortality				X	X	X						

[a] m1–m6 are input-oriented and m7–m9 are output-oriented DEA models.
[b] X indicates that the variable is included into the model.

there is a concern that labor input for countries, which use a large proportion of part-time labor, will be overestimated. The share of the full-time employment in the country is added to control for the difference in working hours across OECD countries.

4. Results and discussion

4.1. Efficiency estimates

Frontier analysis is initially carried out on cross-sectional data for 2007 using a number of alternative models in terms of technology specification and model assumptions. The input and output variables used in the different models are indicated in Table 2.

The DEA efficiency scores corrected for the bootstrap bias and the SFA efficiency scores across different models are presented in Table 3. The reciprocals of efficiency scores in the output-oriented DEA models are used in order to allow for direct comparison with the input-oriented DEA models and the SFA models. The value of 1 indicates that a country produces at the frontier, the lower the value, the farther the country is from the efficient frontier.

The comparison of technical efficiency of the hospital industries across countries should be regarded critically as it concerns the use of resources and does not directly address health outcomes. It is possible for countries, renowned for their good health outcomes in terms of longevity, such as Japan, to perform worse on the technical efficiency side than a country with worse health outcomes, such as Turkey. Countries, which lie on the frontier, are the most successful in their consumption of labor and technology inputs to achieve the amount of hospital discharges adjusted for case severity, relative to other countries.

The Spearman rank correlations are conventionally used to compare efficiency scores provided by various model specifications and make validity judgments [13,46,50,53], although the Pearson correlations between the scores are also sometimes used [52]. The Spearman rank correlation coefficients across the different DEA and SFA models are presented in Table 4.

The Spearman rank correlations between the results of the input-oriented DEA models (m1–m6) obtained using different selections of the input and output variables lie in the range 0.36–0.90. This implies that the change in the definition of hospital employment as well as adding mortality to the output definition might considerably change the efficiency rankings across the OECD sample. On the other hand, the SFA models (s1–s3) provide high Spearman rank correlations (0.89–0.97) between models with different input definitions. The fact that the DEA results are more sensitive to changes in the specification of input and output variables than the SFA models has already been noticed in previous studies [46,52]. Furthermore, the change in the DEA assumptions from input- to output-orientation (the corresponding models m1–m3 and m7–m9) provides consistently high Spearman rank correlations (0.75–0.96) across models with identically defined input and output variables. Thus, while the SFA results suggest good internal validity, the DEA estimations of efficiency are highly sensitive to the choice of variables for inputs and outputs but rather stable to the change in model assumption.

To determine external validity of the DEA and SFA models, the consistency of the results both across time and between the two methodologies is tested. The Spearman rank correlations for the DEA efficiencies between three consecutive years (not presented here) over the time-period 2000–2009 in the model m8 produces correlations between 0.90 and 0.98. High inter-year rank correlations (0.95–0.99) are also observed for the SFA results in the model s2. In terms of the comparability of DEA and SFA, the Spearman rank correlation of 0.43–0.84 is observed between the input-oriented DEA models and the SFA models with identically defined input and output variables. The Spearman rank correlations among the output-oriented DEA models and the SFA models with the identical technology specifications lie in the range 0.74–0.90. The inter-year rank correlations for DEA and SFA suggest that both approaches perform quite well in terms of external validity. The output-oriented DEA models show more consistency with the SFA results than the input-oriented DEA models.

4.2. Regressing efficiency on environmental variables

There are no precise diagnostic tools, which allow choosing the best model specification [52]; however, there is a solid statistical basis for the model selection in case of nested models. Output-orientation in DEA is chosen

Table 3
Efficiency estimation across various model specifications (year 2007).

OECD country[a]	m1	m2	m3	m4	m5	m6	m7	m8	m9	s1	s2	s3
Australia[b,c]	0.95	0.92	.	0.93	0.91	.	0.92	0.92	.	1.00	0.98	.
Austria	0.90	0.84	0.86	0.90	0.84	0.86	0.91	0.90	0.90	1.00	1.00	1.00
Belgium	0.62	0.87	0.87	0.66	0.85	0.87	0.65	0.84	0.86	0.66	0.80	0.76
Canada	0.94	.	.	0.93	.	.	0.65	.	.	0.58	.	.
Czech Republic	0.65	0.61	0.64	0.76	0.60	0.63	0.73	0.68	0.68	0.71	0.70	0.70
Denmark	0.97	0.95	0.93	0.94	0.92	0.91	0.95	0.95	0.93	0.96	0.90	0.92
Estonia	0.85	0.77	0.84	.	.	.	0.85	0.83	0.86	0.88	0.83	0.85
Finland	0.82	0.92	0.86	0.89	0.84	0.87	0.82	0.95	0.86	0.85	0.91	1.00
France	0.91	0.86	0.89	.	.	.	0.91	0.89	0.91	0.94	0.94	0.95
Germany	0.94	0.87	0.89	0.93	0.87	0.89	0.94	0.92	0.92	0.88	0.91	0.90
Greece	0.92	0.83	0.91	.	.	.	0.87	0.87	0.92	1.00	0.85	0.91
Hungary	0.92	0.80	0.79	.	.	.	0.94	0.84	0.83	0.81	0.80	0.82
Iceland	0.60	0.59	0.64	0.62	0.57	0.63	0.68	0.67	0.67	0.65	0.65	0.66
Ireland	0.71	0.52	0.57	0.72	0.55	0.64	0.57	0.57	0.56	0.59	0.59	0.59
Israel	0.97	0.79	0.85	0.95	0.83	0.91	0.94	0.77	0.85	0.82	0.74	0.78
Italy	0.95	0.75	0.78	0.90	0.83	0.92	0.89	0.74	0.79	0.82	0.72	0.75
Japan	.	0.38	0.44	.	0.46	0.46	.	0.40	0.40	.	0.27	0.26
Korea	0.90	0.75	0.87	0.90	0.85	0.87	0.88	0.83	0.86	0.72	0.71	0.73
Luxembourg	0.79	.	.	0.86	.	.	0.80	.	.	0.82	.	.
Netherlands	0.68	0.60	0.69	0.68	0.60	0.74	0.51	0.53	0.53	0.53	0.55	0.55
Norway	0.97	0.95	0.93	0.90	0.86	0.87	0.96	0.96	0.94	0.95	0.91	0.92
Poland	.	0.81	0.83	.	.	.	.	0.85	0.84	.	0.79	0.80
Portugal	0.96	0.64	0.65	0.95	0.88	0.92	0.80	0.60	0.66	0.65	0.57	0.60
Slovak Republic	0.96	0.75	.	0.92	0.74	.	0.94	0.80	.	0.84	0.79	.
Slovenia	.	0.66	0.85	.	0.66	0.88	.	0.74	0.84	.	0.77	0.84
Spain	0.96	0.71	0.72	0.94	0.92	0.92	0.87	0.67	0.73	0.71	0.62	0.66
Switzerland	0.77	0.75	0.78	0.77	0.73	0.77	0.82	0.81	0.80	0.79	0.79	0.80
Turkey	.	0.88	0.87	.	.	.	.	0.85	0.85	.	0.75	0.82
United Kingdom	0.87	.	.	0.86	.	.	0.79	.	.	0.78	.	.
United States	0.94	0.87	0.88	0.92	0.85	0.87	0.87	0.85	0.86	0.92	1.00	1.00
Mean efficiency	0.86	0.77	0.79	0.86	0.77	0.81	0.83	0.79	0.79	0.80	0.77	0.78

[a] New Zealand does not provide data for 2007 but is included in the panel data estimations.
[b] '.' represents a missing value.
[c] Values closer to (farther from) 1 represent more (less) technically efficient countries.

because it might be more appropriate in health production context [13]. Disaggregated hospital employment is a better indication of the specialized labor input across countries than total hospital employment. The selection among the two nested nonparametric models with disaggregated hospital employment, m8 and m9, is performed using the bootstrap method [22,47]. The relevance of the additional input variable in model m9 is tested by calculating the test statistic S under the null hypothesis that the restricted model m8 is correct. The estimate of the one-sided test that $S < 1$ is 0.96, which is not below the critical value from the bootstrap estimation at 5% significance level of 0.936. This suggests that the production technology in model m9 does not provide a significantly better fit of the data, and that m8 is the preferred model. The corresponding selection among the nested SFA models, s2 and s3, is performed using the log-likelihood ratio test. The null hypotheses of the restricted model s2 cannot be rejected, since the test statistics of 1.044 is below the critical Chi-square statistics at the 5% confidence level of 3.84. Thus, both DEA and SFA model specification tests suggest that the inclusion of nurse employment is redundant.

Table 4
The Spearman rank correlation coefficients across various model specifications.

	m1	m2	m3	m4	m5	m6	m7	m8	m9	s1	s2	s3
m1	1.00											
m2	0.40	1.00										
m3	0.46	0.89	1.00									
m4	0.90	0.36	0.37	1.00								
m5	0.62	0.72	0.70	0.70	1.00							
m6	0.80	0.46	0.44	0.82	0.75	1.00						
m7	**0.75**	0.61	0.68	0.64	0.53	0.55	1.00					
m8	0.29	**0.95**	0.88	0.23	0.61	0.29	0.62	1.00				
m9	0.41	0.87	**0.96**	0.33	0.65	0.37	0.71	0.92	1.00			
s1	**0.43**	0.79	0.81	0.39	0.50	0.31	**0.74**	0.84	0.88	1.00		
s2	0.22	**0.84**	0.81	0.20	0.49	0.22	0.55	**0.90**	0.87	0.89	1.00	
s3	0.30	0.83	**0.80**	0.27	0.45	0.28	0.60	0.90	**0.85**	0.90	0.97	1.00

The Spearman rank correlations between the models using the same sets of input and output variables are marked in bold.
The estimation is based on cross-sectional DEA and SFA in the subsample of year 2007 presented in Table 3.

Table 5
Nonparametric (DEA) and parametric (SFA) regression results.

	DEA (m8)			SFA (s2)		
	Estimate	SE		Estimate[a]	SE	
Constant	−0.9447	0.5706	.	−1.5507	0.8296	.
Expenditure	0.0331	0.0091	***	0.1329	0.0233	***
Private expenditure	0.0010	0.0011		0.0027	0.0018	
Inequality	−1.2885	0.2419	***	−1.8560	0.3934	***
Hospital density	0.0041	0.0007	***	0.0074	0.0012	***
Education	0.0079	0.0043	.	0.0101	0.0064	
Length of stay	−0.0196	0.0019	***	−0.0448	0.0030	***
Population over 65	0.0129	0.0042	**	0.0114	0.0062	.
Life expectancy	0.0115	0.0059	.	−0.0030	0.0093	
Infant mortality	0.0353	0.0073	***	0.0406	0.0103	***
Full-time employment	0.0083	0.0017	***	0.0159	0.0027	***
sigma[b]	0.1057	0.0066	***			
sigmaSq[c]				0.0206	0.0028	***
gamma[d]				0.9287	0.0282	***
Log likelihood	199.05			142.337		
Unbalanced panel: cross-sections = 26[e], time periods = 2–10, observations = 202						

Significance level: '***' 0.001; '**' 0.01; '*' 0.05; '.' 0.1.

[a] The coefficients in the SFA model are multiplied with −1 to obtain the effects on efficiency.

[b] sigma is the estimated standard deviation of the assumed left-truncated normal distribution.

[c] sigmaSq is the estimate of the total variance.

[d] gamma is the fraction of the total variance attributable to inefficiency.

[e] Countries in the panel: Australia, Austria, Belgium, Czech Republic, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Japan, Korea, Netherlands, New Zealand, Norway, Poland, Portugal, Slovak Republic, Slovenia, Spain, Switzerland, Turkey, and US.

The panel models are first tested for the presence of time fixed effects. On the basis of log-likelihood ratio tests, the null hypotheses of no time fixed effects cannot be rejected, as both test statistics (13.88 for DEA and 1.98 for SFA) are below the critical Chi-square statistics at the 5% confidence level of 16.92. Therefore, we proceed with estimating the models without year dummies. The results of the two-stage DEA model and the one-step SFA using unbalanced panel data ($T = 2–10$) are listed in Table 5.

The frontier coefficients in SFA, which are as expected positive and significant, are not presented. The positive and significant coefficient for health care expenditure per capita suggests that the hospital sector efficiency rises with the amount of resources invested into the health care system. This finding is consistent with Evans et al. [63], who found that in the sample of 191 countries higher expenditure on health per capita was associated with higher health care efficiency, especially at low expenditure. The coefficient denoting the funds coming from private sources is insignificant, implying that the source of funding is not an important factor determining the hospital sector efficiency. The coefficient for income inequality in a given country is negative and significant, which indicates that a high amount of inequality in a country creates a greater social environmental challenge [12]. The coefficient for average hospital length of stay is negative and significant, which implies that the efficiency of the hospital sector tends to increase with a shorter average length of stay.

With the only exception of life expectancy variable, the signs of the coefficients derived by DEA and SFA are very similar. The difference in DEA and SFA coefficients might be due to the different interpretation of inefficiency by DEA and SFA. While DEA attributes all frontier distance differences between countries to inefficiency, SFA is splitting the variance into an inefficiency component and a random component.

4.3. Sensitivity analyses

Instead of providing the full description of parameter estimates, only the distinguishing features of each analysis are shortly mentioned. First, the original regression of environmental variables on efficiency in DEA and SFA is estimated with health expenditure measured as a share of GDP. The results remain very similar to those above. Then, hospital expenditure instead of total health expenditure is used, which reduces the sample to 23 countries. However, the results are similar to the original estimation, in that increased hospital expenditure in thousands of US$ PPP is associated with higher levels of hospital efficiency. The conclusion is not affected if hospital expenditure is measured as a share of GDP, either. Furthermore, measuring hospital physician employment in full time equivalent (FTE) units in a sample of 18 countries, for which these data are available, does not alter the original conclusions. Finally, when the measure of total hospital employment is used as an input variable in the production function, the main results also remain robust.

5. Conclusion

The paucity of studies comparing hospital efficiency across countries has been underlined by the lack of consistent international data. Although there has been a considerable improvement of the OECD data quality over the last few years, a number of limitations remain. Spinks and Hollingsworth [12] raised concerns to the utilization of the OECD sample as panel data, as most of the series are derived from the moving averages and not obtained as independent samplings. Moreover, the study of the hospital industry is complicated by the inconsistency in reporting standards across countries, the lack of useful control variables, and the fact that data for some countries do not cover all hospitals [56]. More initiatives in future to collect reliable micro and macro data should be encouraged.

The cross-sectional analysis of the hospital sector efficiency in OECD countries using nonparametric and parametric techniques reveals that countries with good health outcomes, such as Japan, can be technically inefficient in their use of health resources. The panel-data analysis shows that countries with more technically efficient hospital sectors tend to have higher health care expenditure per capita. Whether the expenditure is financed through private or public sources is not significant. On the other hand, the hospital sectors in countries with higher income inequality and longer average length of stay are less technically efficient.

However, the aim of this paper was not only to estimate the efficiency of the hospital sector in OECD countries and look at the factors associated with higher efficiency, but also to enhance the understanding of using efficiency methods for international comparison of hospitals. Theoretical considerations suggest that DEA might be more appropriate than SFA in particular cases and vice versa. If both DEA and SFA are deemed appropriate, it may be advisable to check the robustness of the results across the two methodologies [13,50]. Moreover, since the real underlying technology is unknown, alternative specifications in terms of technology and model assumptions might be considered [50]. Further research should target a number of unanswered theoretical questions, particularly, the appropriateness of DEA or SFA in the specific research context and the selection of variables for analysis.

The reliability of parametric and nonparametric efficiency estimations is crucial in decision making from a health policy perspective. Researchers and policy makers should be aware of the limitations and imprecision of parametric and nonparametric techniques in comparing efficiency in the health sector based on aggregate data. Nevertheless, the quantitative international comparison when rigorously conducted might produce a valuable source of evidence for policy.

Acknowledgment

This study was supported by a research grant of the Federal Ministry of Education and Research (BMBF) in Germany (Grant number: 01FL10055). The sponsor had no role in the study design, collection and analysis of data, the writing of the report or the submission of the paper for publication. The authors declare that they have no conflict of interest.

References

[1] OECD. http://www.oecd.org/health/healthpoliciesanddata/oecdhealthdata2012.htm
[2] Dervaux B, Ferrier GD, Leleu H, Valdmanis V. Comparing French and US hospital technologies: a directional input distance function approach. Applied Economics 2004;36(10):1065–81.
[3] Kittelsen SAC, Magnussen J, Anthun K, Häkkinen U, Linna M, Medin E, et al. Hospital productivity and the Norwegian ownership reform—a Nordic comparative study. HERO Working Paper, Oslo University, No. 10; 2008.
[4] Linna M, Häkkinen U, Magnussen J. Comparing hospital cost efficiency between Norway and Finland. Health Policy 2006;77(3): 268–78.
[5] Linna M, Häkkinen U, Peltola M, Magnussen J, Anthun KS, Kittelsen S, et al. Measuring cost efficiency in the Nordic Hospitals—a cross-sectional comparison of public hospitals in 2002. Health Care Management Science 2010;13(4):346–57.
[6] Mobley IV LR, Magnussen J. An international comparison of hospital efficiency: does institutional environment matter? Applied Economics 1998;30(8):1089–100.
[7] Steinmann L, Dittrich G, Karmann A, Zweifel P. Measuring and comparing the (in) efficiency of German and Swiss hospitals. The European Journal of Health Economics 2004;5(3):216–26.
[8] Medin E, Anthun KS, Häkkinen U, Kittelsen SAC, Linna M, Magnussen J, et al. Cost efficiency of university hospitals in the Nordic countries: a cross-country analysis. The European Journal of Health Economics 2011;12(6):509–19.
[9] Färe R, Grosskopf S, Lindgren B, Poullier JP. Productivity growth in health-care delivery. Medical Care 1997;35(4):354.
[10] Hollingsworth B, Wildman J. The efficiency of health production: re-estimating the WHO panel data using parametric and non-parametric approaches to provide additional information. Health Economics 2003;12(6):493–504.
[11] Puig-Junoy J. Measuring health production performance in the OECD. Applied Economics Letters 1998;5(4):255–9.
[12] Retzlaff-Roberts D, Chang CF, Rubin RM. Technical efficiency in the use of health care resources: a comparison of OECD countries. Health Policy 2004;69(1):55–72.
[13] Spinks J, Hollingsworth B. Cross-country comparisons of technical efficiency of health production: a demonstration of pitfalls. Applied Economics 2009;41(4):417–27.
[14] Bhat VN. Institutional arrangements and efficiency of health care delivery systems. The European Journal of Health Economics 2005;6(3):215–22.
[15] Afonso A, Aubyn MS. Non-parametric approaches to education and health efficiency in OECD countries. Journal of Applied Economics 2005;8:227–46.
[16] Alexander CA, Busch G, Stringer K. Implementing and interpreting a data envelopment analysis model to assess the efficiency of health systems in developing countries. IMA Journal of Management Mathematics 2003;14(1):49–63.
[17] Gravelle H, Jacobs R, Jones AM, Street A. Comparing the efficiency of national health systems: a sensitivity analysis of the WHO approach. Applied Health Economics and Health Policy 2003;2(3):141–8.
[18] Grosskopf S, Self S, Zaim O. Estimating the efficiency of the system of healthcare financing in achieving better health. Applied Economics 2006;38(13):1477–88.
[19] Farrell MJ. The measurement of productive efficiency. Journal of the Royal Statistical Society Series A (General) 1957;120(3):253–90.
[20] Jacobs R, Smith PC, Street A. Measuring efficiency in health care. Cambridge: Cambridge University Press; 2006.
[21] Hollingsworth B. The measurement of efficiency and productivity of health care delivery. Health Economics 2008;17(10):1107–28.
[22] Bogetoft P, Otto L. Benchmarking with DEA, SFA, and R. Berlin: Springer Verlag; 2010.
[23] Hollingsworth B, Dawson PJ, Maniadakis N. Efficiency measurement of health care: a review of non-parametric methods and applications. Health Care Management Science 1999;2(3):161–72.
[24] Ozcan YA. Health care benchmarking and performance evaluation: an assessment using data envelopment analysis (DEA). Berlin: Springer Verlag; 2007.
[25] Parkin D, Hollingsworth B. Measuring production efficiency of acute hospitals in Scotland, 1991–94: validity issues in data envelopment analysis. Applied Economics 1997;29(11):1425–33.
[26] Deprins D, Simar L. On Farrell measures of technical efficiency. Recherches Économiques de Louvain/Louvain Economic Review 1983;49(2):123–37.
[27] Fare R, Lovell C. Measuring the technical efficiency of production. Journal of Economic Theory 1978;19(1):150–62.
[28] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision making units. European Journal of Operational Research 1978;2(6):429–44.
[29] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. Management Science 1984;30:1078–92.

[30] Hollingsworth B, Smith P. Use of ratios in data envelopment analysis. Applied Economics Letters 2003;10(11):733–5.
[31] Simar L, Wilson PW. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. Management Science 1998;44:49–61.
[32] Simar L, Wilson PW. A general methodology for bootstrapping in non-parametric frontier models. Journal of Applied Statistics 2000;27(6):779–802.
[33] Simar L, Wilson PW. Estimation and inference in two-stage, semi-parametric models of production processes. Journal of Econometrics 2007;136(1):31–64.
[34] Aigner D, Lovell C, Schmidt P. Formulation and estimation of stochastic frontier production function models. Journal of Econometrics 1977;6(1):21–37.
[35] Meeusen W, van Den Broeck J. Efficiency estimation from Cobb–Douglas production functions with composed error. International Economic Review 1977;18(2):435–44.
[36] Ruggiero J. Efficiency estimation and error decomposition in the stochastic frontier model: a Monte Carlo analysis. European Journal of Operational Research 1999;115(3):555–63.
[37] Rosko MD, Mutter RL. Stochastic frontier analysis of hospital inefficiency: a review of empirical issues and an assessment of robustness. Medical Care Research and Review 2008;65(2):131–66.
[38] Battese GE, Coelli TJ. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. Journal of Econometrics 1988;38(3):387–99.
[39] Kumbhakar SC. Production frontiers and panel data: an application to US class 1 railroads. Journal of Business & Economic Statistics 1987;5:249–55.
[40] Pitt MM, Lee LF. The measurement and sources of technical inefficiency in the Indonesian weaving industry. Journal of Development Economics 1981;9(1):43–64.
[41] Schmidt P, Sickles RC. Production frontiers and panel data. Journal of Business & Economic Statistics 1984;2(4):367–74.
[42] Battese GE, Coelli TJ. Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India. Journal of Productivity Analysis 1992;3(1):153–69.
[43] Kumbhakar SC. Production frontiers, panel data, and time-varying technical inefficiency. Journal of Econometrics 1990;46(1–2):201–11.
[44] Lee YH, Schmidt P. A production frontier model with flexible temporal variation in technical efficiency. In: Fried HO, Lovell CAK, Schmidt SS, editors. The measurement of productive efficiency: techniques and applications. New York: Oxford University Press; 1993. p. 237–55.
[45] Battese GE, Coelli TJ. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 1995;20(2):325–32.
[46] Webster R, Kennedy S, Johnson L. Comparing techniques for measuring the efficiency and productivity of Australian private hospitals. Working Papers in Econometrics and Applied Statistics, No. 98/3; 1998.
[47] Simar L, Wilson PW. Testing restrictions in nonparametric efficiency models. Communications in Statistics-Simulation and Computation 2001;30(1):159–84.
[48] Simar L, Wilson PW. Non-parametric tests of returns to scale. European Journal of Operational Research 2002;139(1):115–32.
[49] Valdmanis V. Sensitivity analysis for DEA models: an empirical example using public vs. NFP hospitals. Journal of Public Economics 1992;48(2):185–205.
[50] Giuffrida A, Gravelle H. Measuring performance in primary care: econometric analysis and DEA. Applied Economics 2001;33(2):163–75.
[51] Chirikos TN, Sear AM. Measuring hospital efficiency: a comparison of two approaches. Health Services Research 2000;34(6):1389–408.
[52] Jacobs R. Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. Health Care Management Science 2001;4(2):103–15.
[53] Linna M. Measuring hospital cost efficiency with panel data models. Health Economics 1998;7(5):415–27.
[54] Bogetoft P, Otto L. Benchmark package. Technical Report, R; 2012.
[55] Coelli T, Henningsen A. Package 'frontier'. Technical Report, R; 2012.
[56] OECD. Health at a Glance 2011: OECD indicators. OECD Publishing; 2011, http://dx.doi.org/10.1787/health_glance-2011-en.
[57] Ozcan YA, Luke RD, Haksever C. Ownership and organizational performance: a comparison of technical efficiency across hospital types. Medical Care 1992;30(9):781–94.
[58] Gerdtham UG, Löthgren M, Tambour M, Rehnberg C. Internal markets and health care efficiency: a multiple-output stochastic frontier analysis. Health Economics 1999;8(2):151–64.
[59] OECD. OECD Health Data 2012. Definitions, sources and methods. International Shortlist for Hospital Morbidity Tabulation (ISHMT); 2012. http://stats.oecd.org/wbos/fileview2.aspx?IDFile=e477970b-3024-4188-8dc6-13f3db201846
[60] Herr A. Cost and technical efficiency of German hospitals: does ownership matter? Health Economics 2008;17(9):1057–71.
[61] Tiemann O, Schreyögg J. Effects of ownership on hospital efficiency in Germany. Business Research 2009;2(2):115–45.
[62] Tiemann O, Schreyögg J. Changes in hospital efficiency after privatization. Health Care Management Science 2012;15(4):1–17.
[63] Evans DB, Tandon A, Murray CJL, Lauer JA. Comparative efficiency of national health systems: cross national econometric analysis. British Medical Journal 2001;323(7308):307–10.
[64] Economou C, Giorno C. Improving the performance of the public health care system in Greece. OECD Economics Department Working Papers, O Publishing; 2009.
[65] Pampel FC, Pillai VK. Patterns and determinants of infant mortality in developed nations, 1950–1975. Demography 1986;23(4):525–42.
[66] Wennemo I. Infant mortality, public policy and inequality—a comparison of 18 industrialised countries 1950–85. Sociology of Health & Illness 1993;15(4):429–46.
[67] Kawachi I, Kennedy BP, Lochner K, Prothrow-Stith D. Social capital, income inequality, and mortality. American Journal of Public Health 1997;87(9):1491–8.
[68] DeFelice LC, Bradford WD. Relative inefficiencies in production between solo and group practice physicians. Health Economics 1997;6(5):455–65.
[69] Or Z. Determinants of health outcomes in industrialised countries: a pooled, cross-country, time-series analysis. OECD Economic Studies 2000;30:53–78.
[70] Norton EC, Van Houtven CH, Lindrooth RC, Normand SLT, Dickey B. Does prospective payment reduce inpatient length of stay? Health Economics 2002;11(5):377–87.
[71] Herr A, Schmitz H, Augurzky B. Profit efficiency and ownership of German hospitals. Health Economics 2011;20(6):660–74.
[72] Street A. How much confidence should we place in efficiency estimates? Health Economics 2003;12(11):895–907.